On Iterative Neural Network Pruning, Reinitialization, and the Similarity of Masks

Michela Paganini (Facebook AI Research), Jessica Zosa Forde (Brown University)

FACEBOOK

In the context of the Lottery Ticket Hypothesis (Frankle & Carbin, 2018)...

Do different pruning methods identify the same "winning ticket"? Or is there more than one?

Research Findings

1. ∃ multiple different lottery tickets within an over-parametrized network

2. possible to find them through a variety of pruning techniques *[extends the findings of Zhou et al. (2019)]*

3. random structured pruning > random unstructured pruning
   ➡ nets are more resistant to the removal of random units/ channels than random individual connections
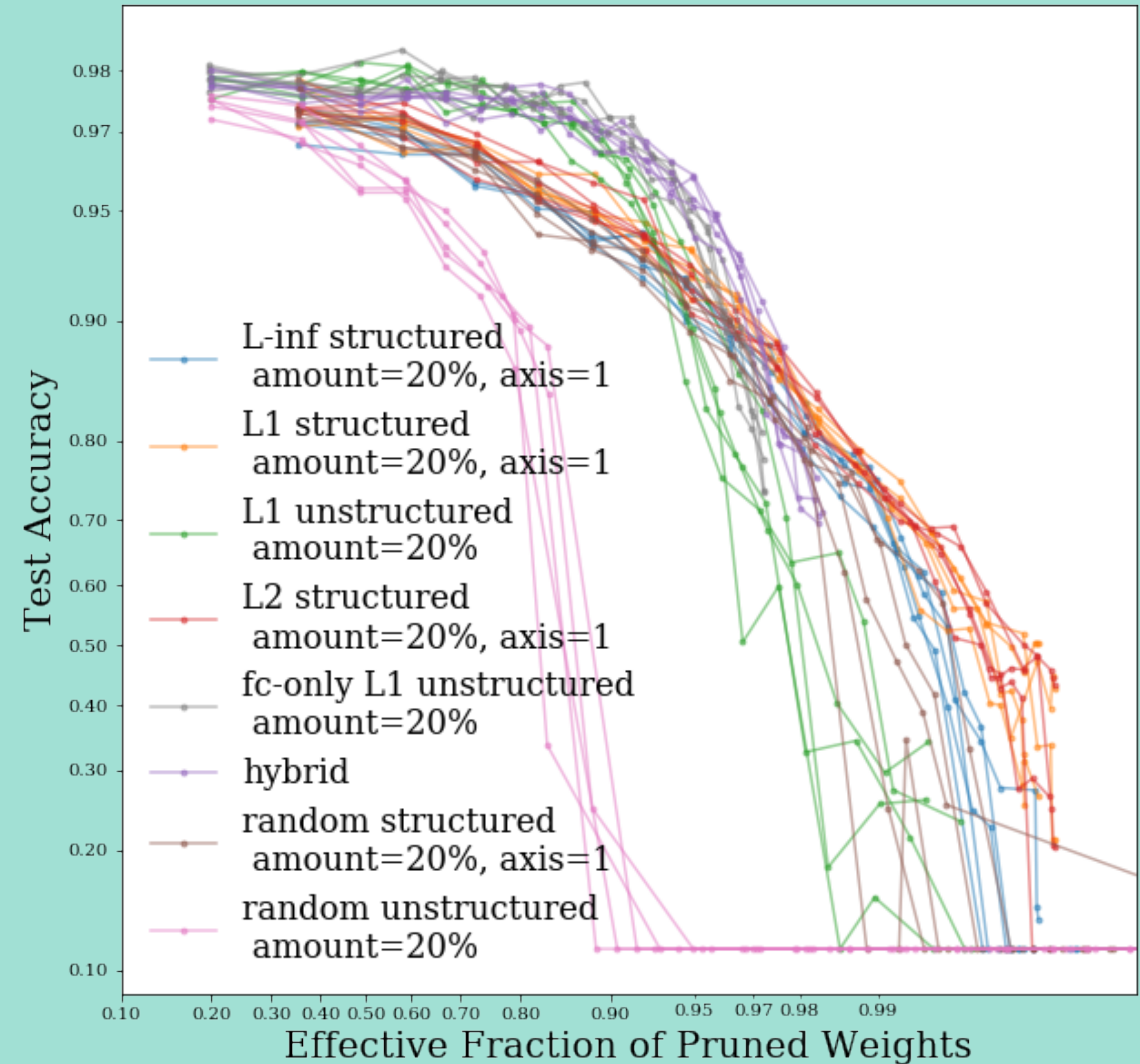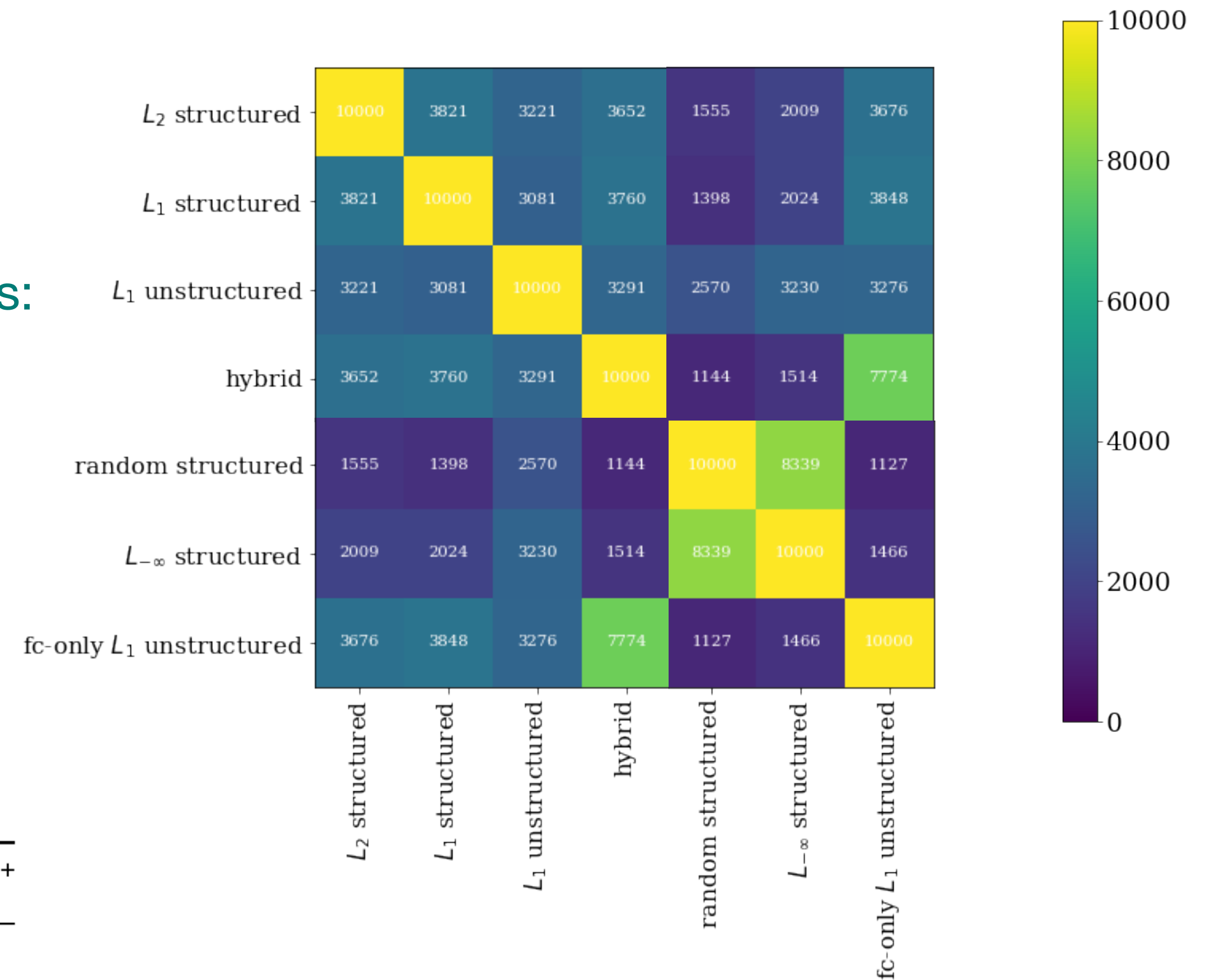
Research Findings

5. networks pruned via different iterative pruning techniques learn vastly different functions of their input, and similarly performing networks make different mistakes on held-out test sets

Number of examples in the MNIST test set over which the sub-networks obtained through each pruning technique agree on the prediction, on average (over 5 seeds), after 18 pruning iterations:


Sub-network accuracies at each pruning iteration + ensembling:

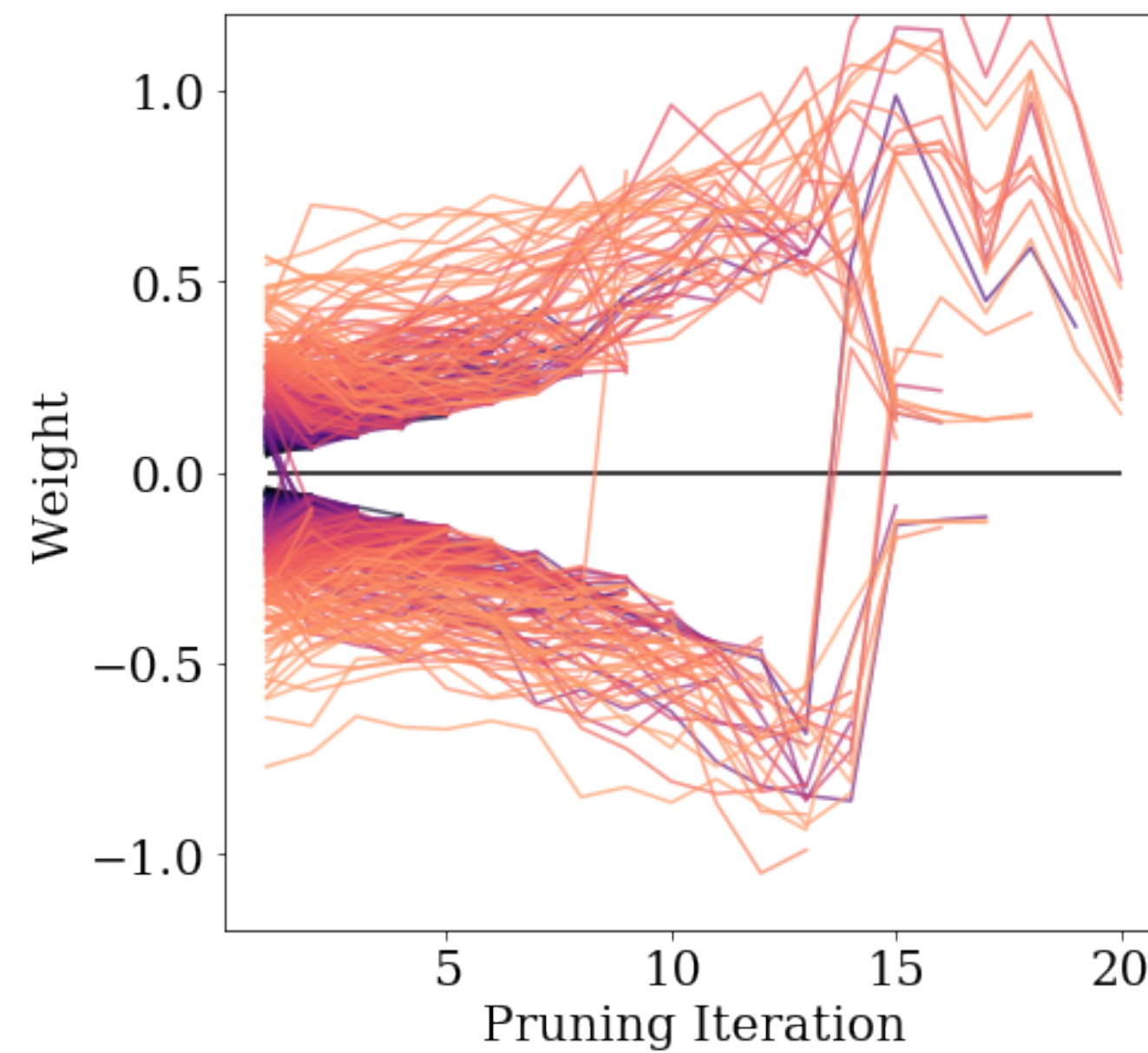| Pruning Iteration	L_2 S	L_1 S	L_1 US	hybrid	random S	$L_{-\infty}$ S	fc-only L_1 US	all	hybrid + fc-only + L_1 US
18	36.7	37.8	32.4	81.8	11.3	14.5	87.4	91.0	91.6

Research Findings

6. weight stability to pruning correlates with performance, and can be induced through suitable pruning techniques, even without late resetting
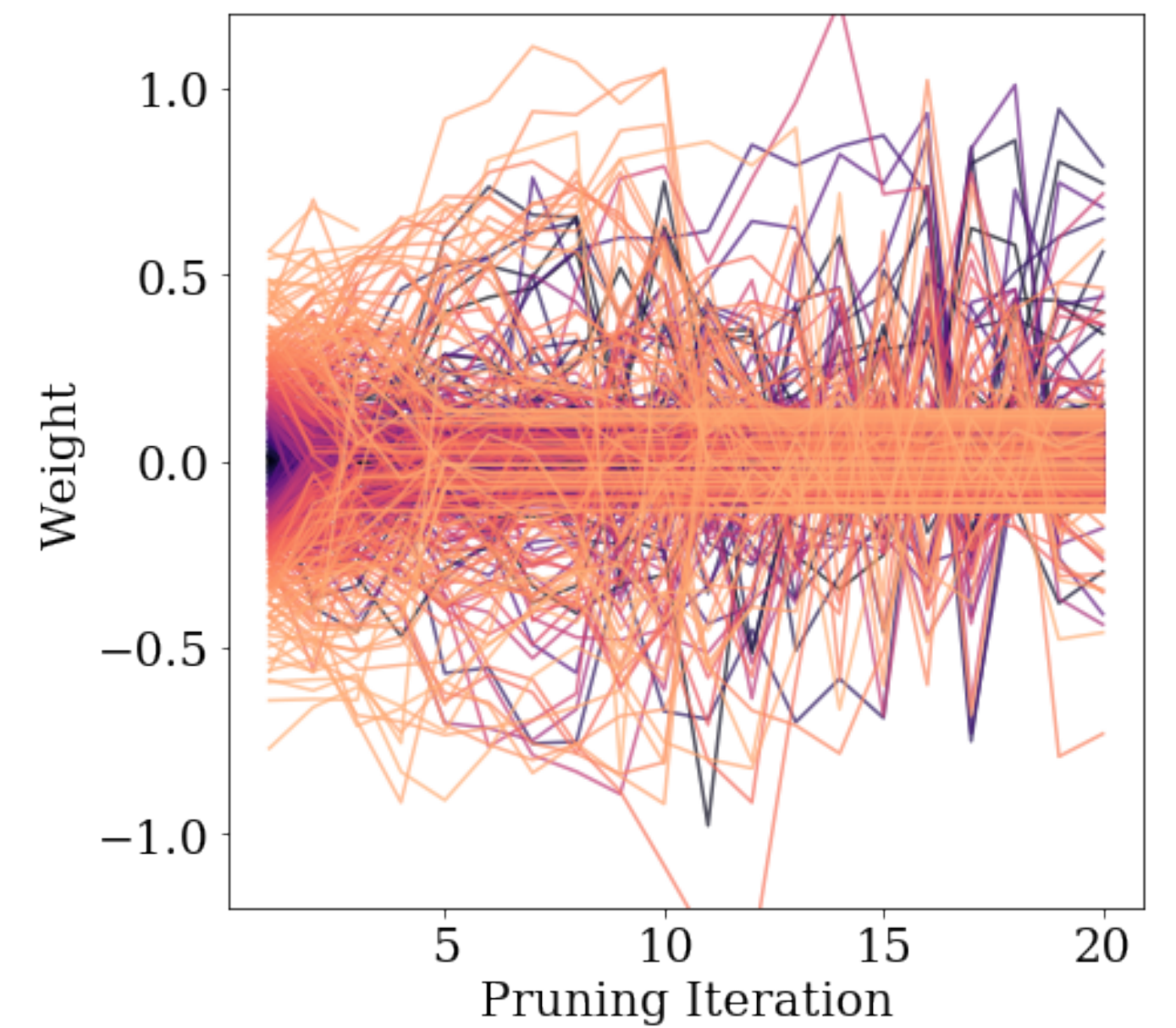
Weight values (y-axis) after 30 epochs of training at various consecutive sparsity levels (x-axis), for weights in the 2nd convolutional layer in the LeNet architecture (seed: 0)
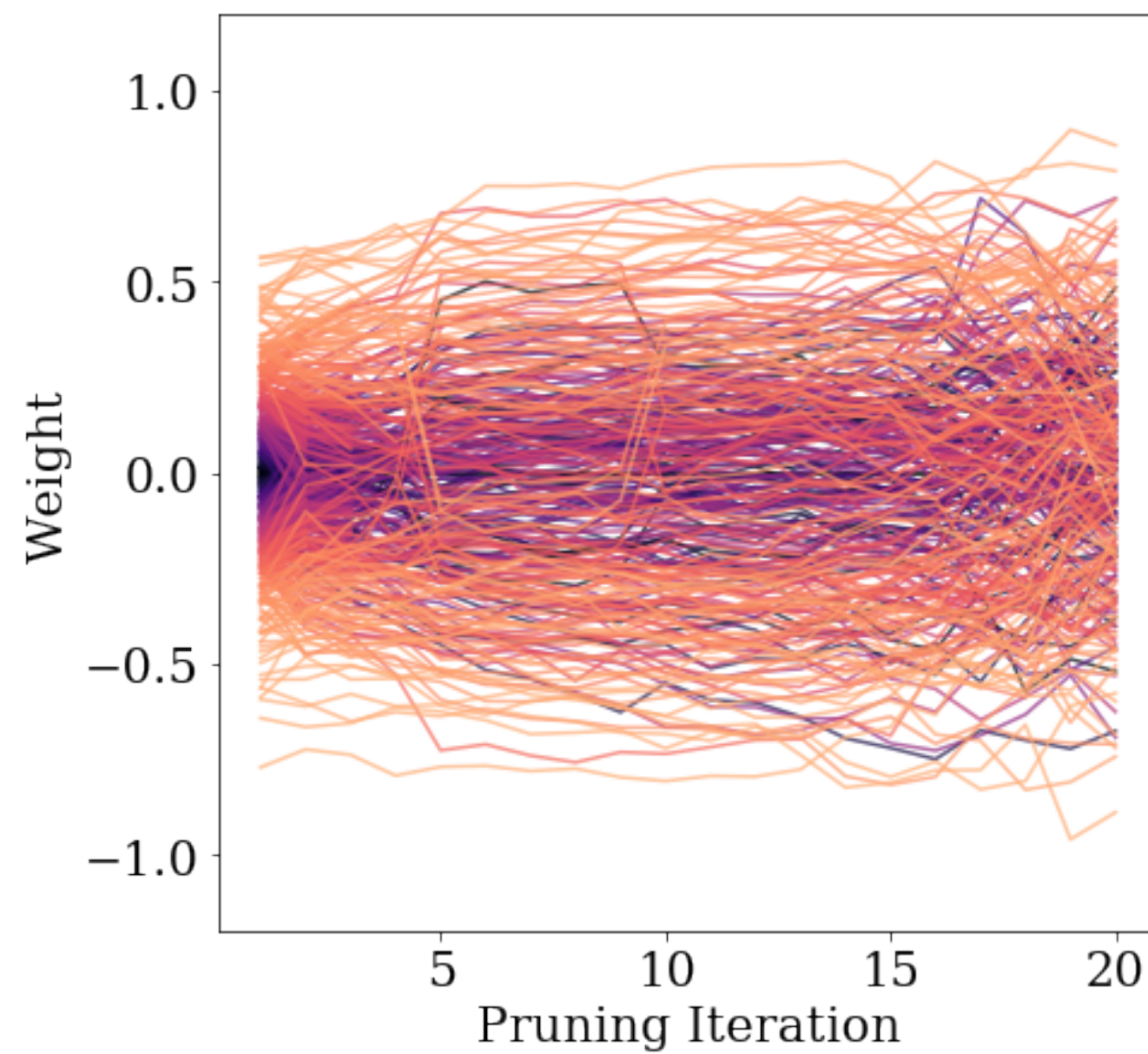
hybrid structured (conv) and unstructured (FC)　　　　**unstructured**　　　　**structured**

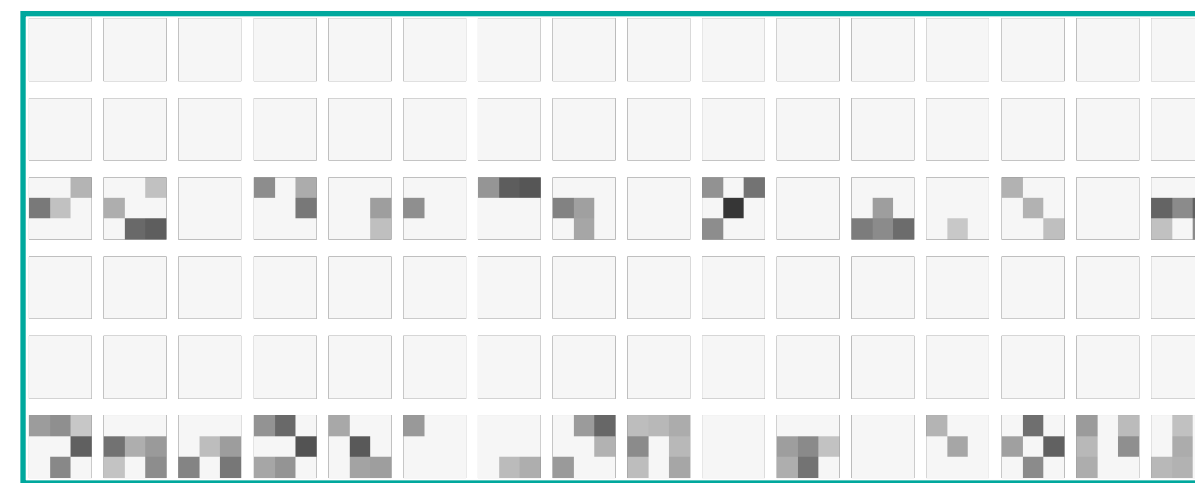# Should I rewind or fine-tune?
What's the difference?

Research Findings

4. lottery ticket-style weight rewinding, coupled with unstructured pruning, gives rise to connectivity patterns similar to structured pruning (~feature selection). Not true for finetuning.
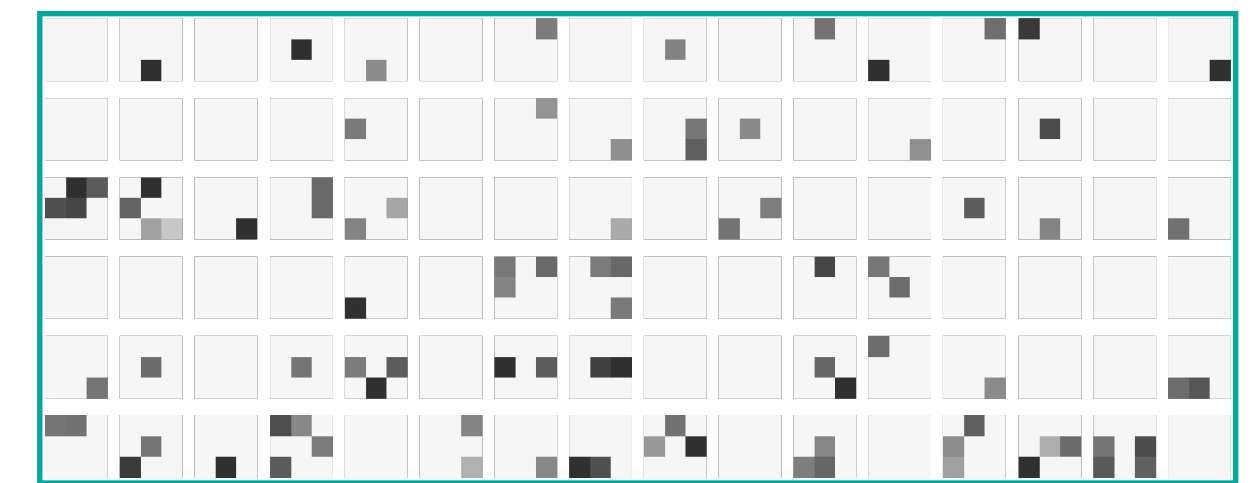
LeNet conv1 weights
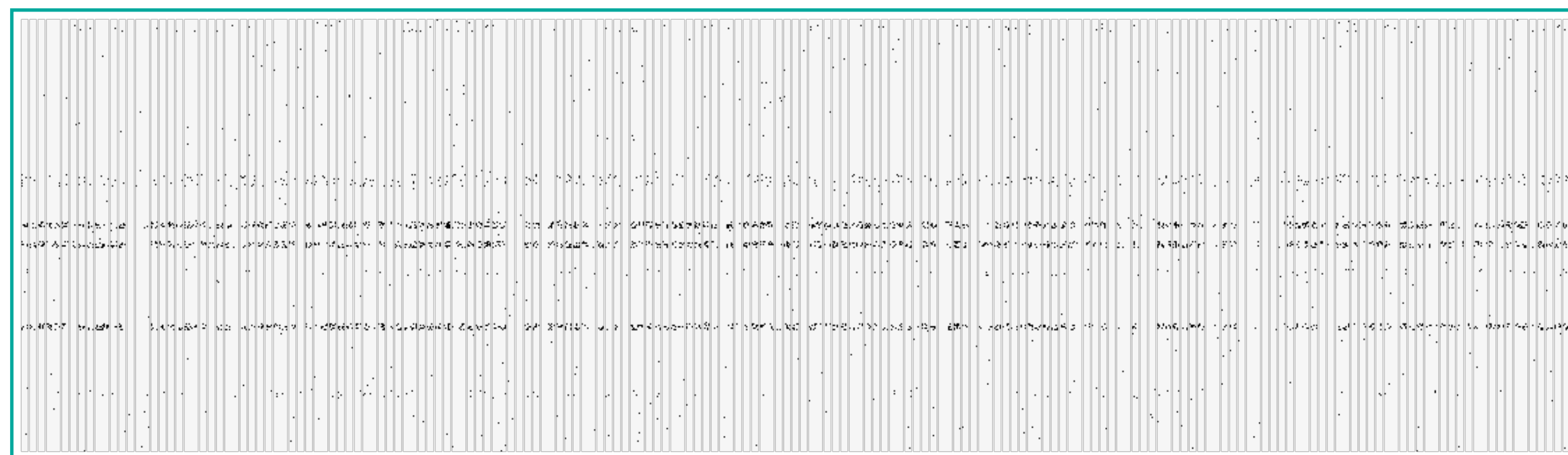

structured pruning + rewinding
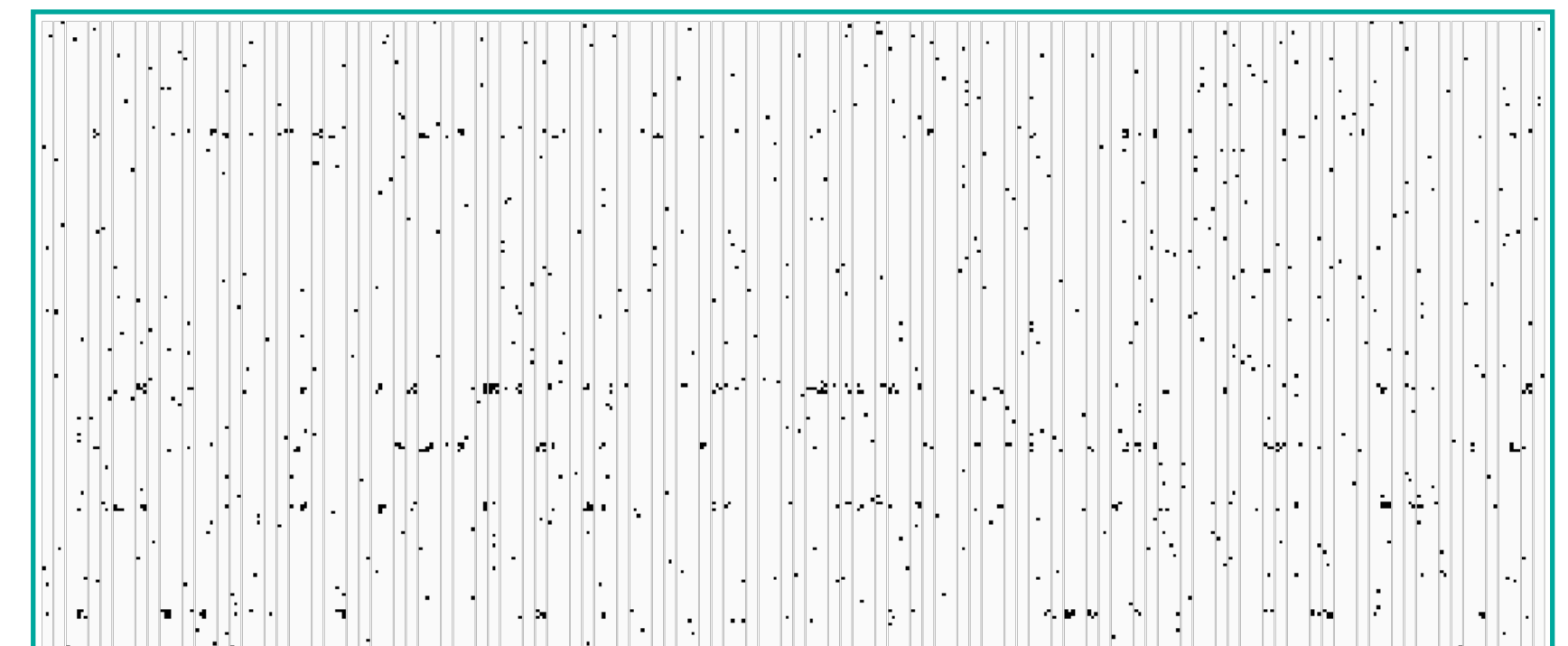

unstructured pruning + rewinding

unstructured pruning + finetuning

AlexNet conv2 weights

unstructured pruning + rewinding

VGG11 conv2 weights

unstructured pruning + rewinding

Questions?

Contact us: michela@fb.com, jessica_forde@brown.edu